FEBS Letters 584 (2010) 2556–2564

journal homepage: www.FEBSLetters.org

Review

Use of genome-scale metabolic models for understanding microbial physiology

Liming Liu, Rasmus Agren, Sergio Bordel, Jens Nielsen *

Department of Chemical and Biological Engineering, Chalmers University of Technology, Kemivägen 10, SE 412 96 Gothenburg, Sweden

ARTICLE INFO

Article history:
Received 14 April 2010
Revised 18 April 2010
Accepted 20 April 2010
Available online 24 April 2010

Edited by Stefan Hohmann

Keywords:
Genome-scale metabolic model
Reconstruction
Algorithm
Microbial physiology

ABSTRACT

The exploitation of microorganisms in industrial, medical, food and environmental biotechnology requires a comprehensive understanding of their physiology. The availability of genome sequences and accumulation of high-throughput data allows gaining understanding of microbial physiology at the systems level, and genome-scale metabolic models represent a valuable framework for integrative analysis of metabolism of microorganisms. Genome-scale metabolic models are reconstructed based on a combination of genome sequence information and detailed biochemical information, and these reconstructed models can be used for analyzing and simulating the operation of metabolism in response to different stimuli. Here we discuss the requirement for having detailed physiological insight in order to exploit microorganisms for production of fuels, chemicals and pharmaceuticals. We further describe the reconstruction process of genome-scale metabolic models and different algorithms that can be used to apply these models to gain improved insight into microbial physiology.

1. Introduction

1.1. The challenge of microbial fermentations: comprehensive understanding of the microbial physiology

Microorganisms are able to produce a wide range of valuable compounds and microbial fermentation is widely used as a route for production of chemicals in industry. Microbial fermentation is expected to increase to about 20% of the total global chemical market [1], which will represent a sales volume of US$310 billion. This increase is expected in basic chemicals and commodities as well as in the field of polymers and high added value fine chemicals. Furthermore, microbial fermentation is making novel products and specialty chemicals available which have not previously been synthesized by traditional petrochemical processes. The efficiency of a microbial production process should be evaluated based on the following key elements: high titer, high yield, high productivity and process robustness. The mentioned parameters highly depend on the host microorganism. Natural screening, mutagenesis, selection, bioprocess development, genetic engineering and metabolic engineering strategies have been adopted to increase the metabolic capabilities of the host microorganisms [2]. Improved strains have been proven to be commercially successful in the production of organic acids [3], amino acids [4], biofuels [5], and pharmaceuticals [6]. Nevertheless, problems such as the accumulation of toxic intermediates or metabolic stress resulting in a decreased cellular fitness are still far from being solved. The lack of knowledge about the regulatory mechanisms of key enzymes and the complex relationships between microbial genotype and metabolic phenotype are still barriers to the development of efficient cell factories. Over-expression, deletion or introduction of heterologous genes in specific metabolic pathways does not always result in the desired phenotype. A good example is attempts to increase the glycolytic flux, which cannot be increased by individual or combinational over-expression of genes encoding the key glycolytic enzymes in either eukaryotic or prokaryotic microorganisms [7]. The essence of the problems listed above lies in the lack of a comprehensive and global understanding of the microbial physiology. Microbial physiology is the result of the interplay between gene transcription, protein translation, enzyme kinetics, internal composition (e.g. the intracellular energy charge and the interior redox potential and intracellular pH) [8,9], and extracellular environment [10,11], such as temperature, dissolved oxygen and other sources of physical or chemical stress. The microbial genotype, represented by the genome, is typically evolved for survival and growth in their natural habitat, and it is often necessary to retrofit the genotype to obtain a desired phenotype [12]. Random mutagenesis and metabolic engineering strategies have been used to either modify or introduce new cellular metabolic capabilities [13]. These strategies include relief of feedback inhibition, deletion of competing pathways, up-regulation of primary biosynthetic pathways, over-expression of export processes, and insertion of new metabolic pathways.

The total protein and RNA concentration in bacteria such as *Escherichia coli* is in the range 300–400 g/L bacteria volume [14],

* Corresponding author. Fax: +46 31 772 3801.
E-mail address: nielsenj@chalmers.se (J. Nielsen).

0014-5793/$36.00
doi:10.1016/j.febslet.2010.04.052

Table 1
The reactions and enzymes number involved in nucleotides in KEGG [40].

	ATP	ADP	NADH	NAD^+	NADPH	$NADP^+$	CoA	Acetyl-CoA
Reactions number	496	347	740	750	887	889	480	169
Enzymes number	454	350	433	455	462	462	250	119

and in general microbial cells are 20–30% volume-occupied by macromolecules. If the total concentration of these macromolecules inside cells accumulated to a significant proportion of the cellular volume, this would lead to macromolecular crowding [14], which affects the folding and aggregation of proteins and the metabolic capability of microbial cells. Small metabolites also play complex roles in metabolic regulation. It has been well-documented that nucleotides not only make up the structural units of RNA and DNA, but also play central roles in metabolism [15]. Nucleotides serve as energy sources [16], co-factors for enzymatic reactions (Table 1 and Fig. 1) or for transport reactions [17,18], and participate in protein transcription and translation [19], cellular signaling, cell morphology [20] and stress response.

Microarray analysis has demonstrated that sets of functionally related genes are coordinately induced or repressed in response to developmental or environmental cues [21]. Therefore, the microbial metabolic function is subject to hierarchical control [22] and originates at the level of transcription (induction–repression mechanism and mRNA degradation), translation (protein activation and proteolysis) and enzyme activity (allostery) or usually a combination of them (such as signaling cascades). Manipulation of the gene regulation mechanisms has been developed as a rational strategy to obtain the desired phenotypes based on careful analysis of the regulation control mechanisms, and successfully applied in the control of carbohydrate utilization [23,24], lipid metabolism [25], amino-acid starvation response [26], nutrient and environmental stress response [27] and expression of the ribosomal regulon [28]. While traditional metabolic engineering strategies are based predominantly on single or multiple genetic manipulations of the target metabolic pathway, many complex aspects of metabolism cannot be altered as desired by manipulating particular metabolic genes and a more holistic view becomes necessary. With the accumulation of genome sequences and other genome scale data, reconstruction of metabolic, regulatory, and signaling networks has become possible, which allows for a complete and comprehensive understanding of the microbial cell as a whole.

1.2. The accumulation of high-throughput data

According to http://www.ebi.ac.uk/, there are 1051 microbial genome sequences that have been completed, and the sequencing of other 2527 microbial genomes is in progress. Another 1455 microbial species have been partially sequenced. The number of complete genome sequences continues to increase and new information generated from the already completed sequences is also exploding. As illustrated in Fig. 2, the amount of high-throughput data (transcriptome, proteome, enzyme, metabolites, biochemical and binary protein reaction, and biological network) originating from genome sequences is rapidly increasing and is at the point of becoming a flood of biological information. Another important data source is the scientific literature, which records an overwhelming amount of biological facts. With MEDLINE as example, MEDLINE contains 11 million references to journal articles in the life sciences, and the size is continuously growing at a rate of several thousand papers per week. The huge and increasing textual resources provide the scientists with (1) the name, sequence, location and transcriptional regulation, plus the essentiality of genes for cell growth; (2) the sequence, domain, motif, structure, and function of proteins [29]; (3) nomenclature (enzyme name, EC number, etc.), functional parameters, organism-related information, enzyme structure, stability and application of an enzyme [30]; (4) metabolites, enzyme inhibitor and activator, co-factors of metabolic reactions and pathways [31]; (5) metabolic and protein interaction networks [32]; (6) transcriptional regulation, membrane transporters, and metabolism; and (7) the composition, and the physiological parameters of strains [33]. The platforms for the literature include Pubmed, MEDLINE, and CiteXplore. The latter combines literature search with text mining tools for biology and it contains the papers from Agricola (506447 citations), Chinese biological abstracts (134890 citations), Citeseer (716328 citations), patents (1902475 citations), MEDLINE (19474348 citations), and Pubmed (493277 citations). A series of techniques have been developed to identify, extract, manage, integrate and exploit this knowledge, and to discover new, hidden or unsuspected information, such as Textpresso [34], PubFinder [35], PubMatrix [36], WIKIGENE [37], and MineBlast [38]. Based on the results from literature mining, systems biology aims to connect the textual evidences to biochemical and signaling models in order to produce accurate computational models. Genome-scale metabolic models are important platforms that combine wet (high-throughput data) and dry (bibliographic data) data to decipher the metabolic, regulatory, and signaling networks at the level of the whole organism.

2. Reconstruction of genome-scale metabolic models

The reconstruction process of a genome-scale metabolic model involves the following five steps [39]: (1) creation of a draft model; (2) reconstruction of a detailed model; (3) conversion into a mathematical format; (4) identification and filling of gaps; and (5) simulation and visualization. These steps are described in detail in Fig. 3.

The reconstruction process starts with the identification of coding sequences and the functional annotation of genes, which can be accomplished by sequence homology with known proteins. Several databases provide such functional annotations for most of the sequenced microbial genomes. It should be noted that only genes encoding for enzymes or membrane transporters are used for model reconstruction, not the complete set of sequenced genes. In order to build a metabolic model, it is necessary to translate the functional annotation of enzymes into appropriate biochemical reactions (Fig. 3A). EC numbers provide an unambiguous way to identify enzyme-catalyzed reactions, KEGG and other databases provide direct correspondence between EC numbers and reactions. However, there are several special cases that should be considered: (1) recently identified enzymatic activities, which have not yet been assigned an EC number; (2) annotation which only provides more general GO categories and not full EC numbers; (3) some enzymes have a broad specificity and accept multiple substrates, one example being alcohol dehydrogenase (EC1.1.1.1), which catalyzes the oxidation of several distinct primary alcohols; (4) isoenzymes are encoded by different genes, but catalyze the same reaction; and (5) there are several genes associated with one or more reactions catalyzed by an enzyme complex. To address these issues, a number of software packages (i.e. PATHOLOGIC) for textual annotation in UniProtKB/SwissProt/BRENDA and literature mining have to be used to identify gene–reaction associations.

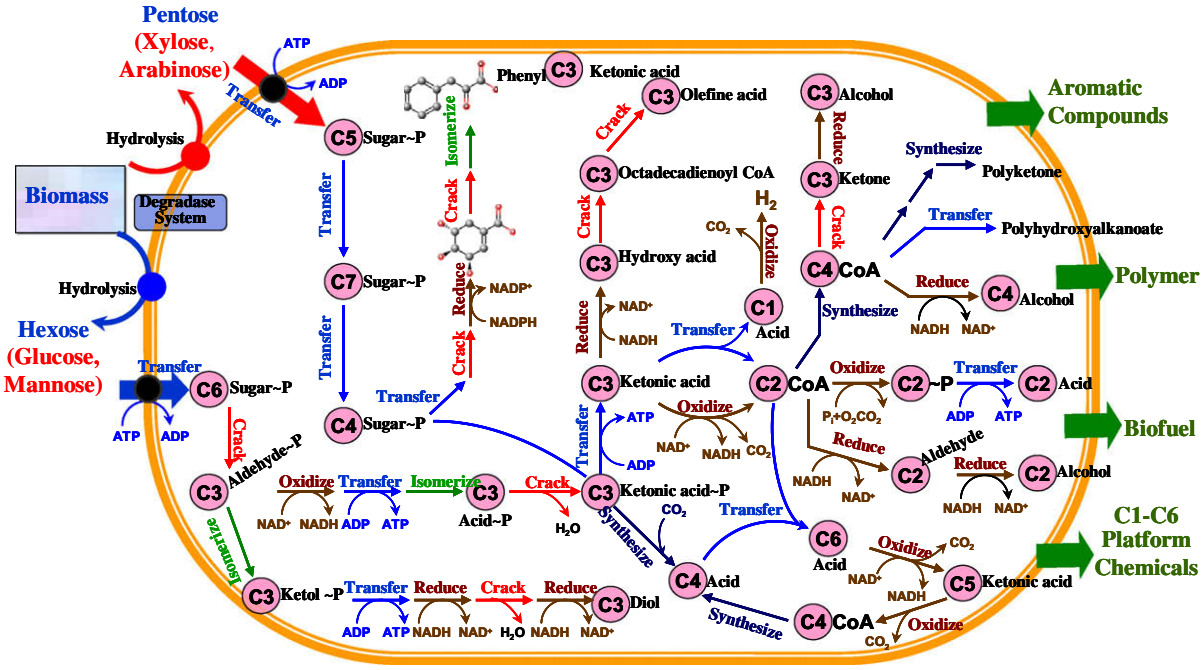

Fig. 1. An illustration of the wide use of the co-factor couple NADH/NAD$^+$ in different metabolic reactions involved in the production of key metabolic products identified as platform chemicals.

To obtain candidate biochemical reactions, the identified genes with assigned EC numbers are matched to biochemical reaction databases, e.g. KEGG [40] and BREAD [41], or to already existing metabolic models for phylogenetically close organisms (e.g. the models in BIGG database (http://bigg.ucsd.edu/)). The gene names and details about individual reactions, such as (1) metabolites and co-factors; (2) chemical formula or InChI string for each metabolite; (3) identifiers for each metabolite from compound databases such as CHEBI (http//www.ebi.ac.uk/chebi), PubChem (http://ncbi.nlm.nih.gov/sites/entrez), or KEGG (http://www.genome.jp); (4) reaction stoichiometry and directionality; (5) gene and reaction localization; (6) reaction identifier from reaction databases such as KEGG; and (7) metabolic subsystems should all be included in the reaction annotation. It should be noted that this stage only collects candidate biochemical reactions, and that many proteins which catalyze DNA methylation, rRNA modification, or function as kinases in signal transduction reactions also have EC numbers but they should not be included in this reaction list.

A crucial step in the reconstruction of genome-scale metabolic models is the addition of reactions which are not inferred from genome annotation (Fig. 3B). Such reactions include: (1) spontaneous reactions; (2) extracellular transport reactions; (3) intracellular transport reactions; and (4) exchange reactions. The addition of these reactions will decrease the number of dead-end metabolites and improve network connectivity.

Another set of required reactions are the biomass reactions, which describe all known biomass constituents and their contribution to the overall cellular biomass. The biomass reaction stoichiometry is determined by the relative amounts of necessary macromolecules and building blocks in the biomass composition of the target organism, and is usually derived from the literature studies. If such biochemical and physiological data are not available, this has to be determined experimentally for cells growing in the logarithmic phase. Information about energy requirement (in terms of ATP molecules) for microbial cell growth (e.g. proteins, DNA and RNA synthesis) and maintenance (e.g. turgor pressure) is also included in the biomass equation. The values for these energy requirements can be found in the literature or be determined from chemostat growth experiments. The reconstruction needs to contain information regarding growth-enabling media, including: (1) the available substrates; (2) the auxotrophic components; (3) minimal medium composition; (4) rich medium composition; and (5) uptake or secretion rates.

As shown in Fig. 3C, the third stage of the reconstruction process is to convert the reaction list into a constraint-based mathematical format. If no constraints are imposed on the model it can take up or excrete compounds at any arbitrary rate, which is clearly not representative of living cells. Therefore, constraints have to be applied in a manner so that the flux solution space better approximates that of a living organism. These constraints are normally set to represent known growth behavior. There are four different kinds of constraints that limit the cellular functions: physiochemical constraints, topological constraints, environmental conditions constraints, and regulatory constraints. Genome-scale metabolic models normally rely on physiochemical and/or environmental conditions constraints: flux balance ($S \cdot v = 0$), energy balance ($\Delta E = 0$), enzyme or transporter capacity ($v_i \leqslant v_{max}$) and thermodynamics ($0 \leqslant V_{min}$).

Once a draft metabolic model is obtained, the consequent steps are to verify and evaluate the reconstructed model (Fig. 3D). The first step is to check the consistency of the reconstructed model, to identify metabolic gaps and to track how different substrates are catabolized into different metabolites. An intensive textual mining in biochemical databases, biochemical textbooks, and the published literature will help to fill possible gaps. However, it should be noted that adding new reactions to the network may cause new gaps. Therefore, when adding reactions, it should be ensured that all metabolites are connected to the network.

The following steps in the reconstruction will be testing of the models' metabolic capacity at three distinct levels: (1) test the ability to produce biomass components in different media and under all known growth conditions. The aim of this step is to validate the capabilities of the reconstructed network; (2) test the growth ability and rate on different types of media; and (3) test whether the model can produce the secreted metabolites from a given

Fig. 2. Overview of different types of high-throughput data that are available in biochemical databases.

substrate and whether the specific production rates of metabolites are correct. It is likely that this step may lead to the requirement for further addition of reactions with or without experimental evidence. The last step of the reconstruction process is to use the model to predict growth phenotypes. By comparison to experimental data, model predictions will help in further refining the model (Fig. 3D).

3. Algorithms help understanding microbial physiology

A genome-scale metabolic model is basically just a stoichiometric representation of all possible metabolic reactions in the cell. The stoichiometry and the reversibility of each reaction, together with the assumption of steady state for the internal metabolites, allow defining a region of allowed flux distributions. Several

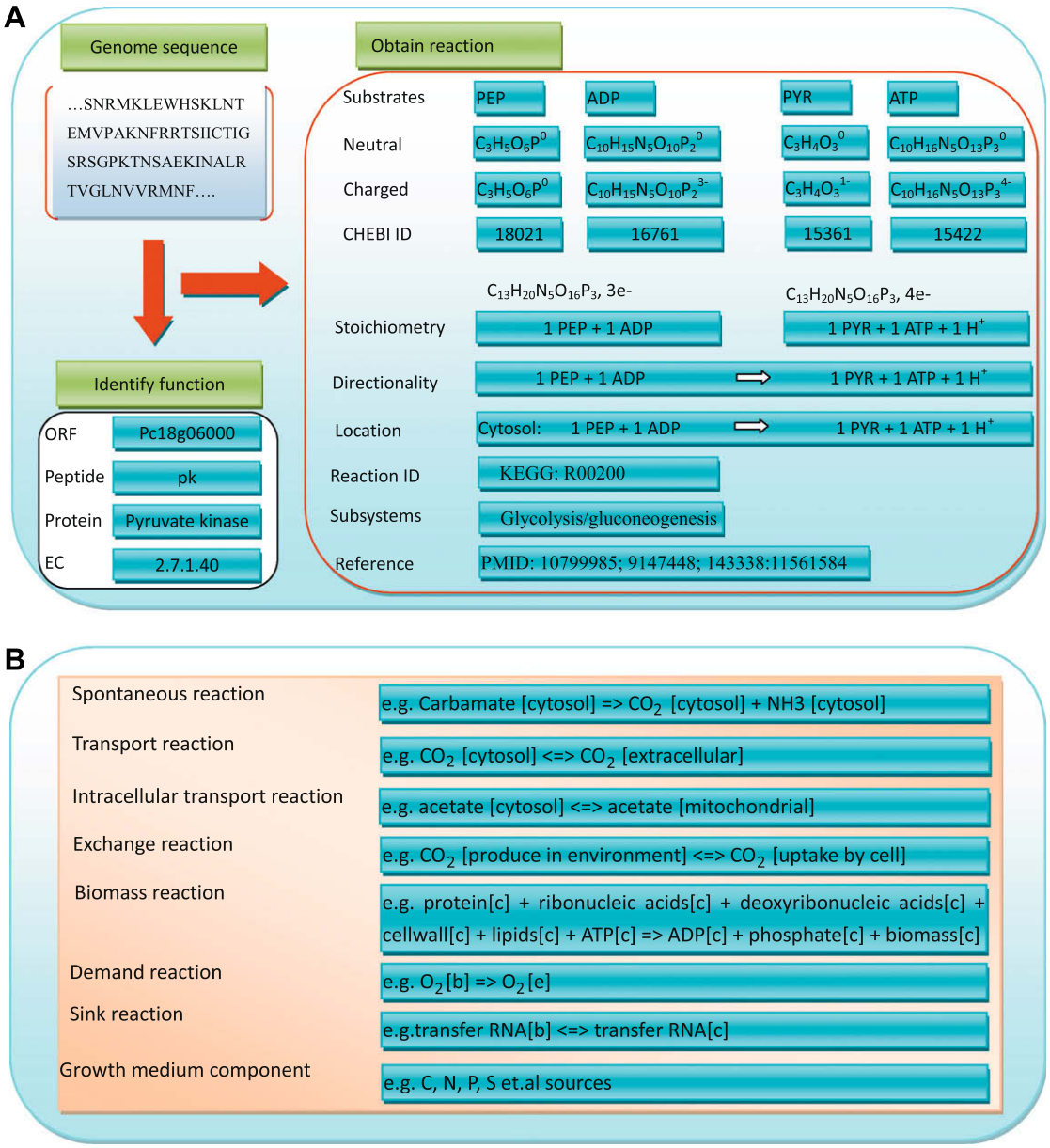

Fig. 3. Description of the different steps in the reconstruction of genome-scale metabolic models.

elegant algorithms have been developed in order to use genome scale models to explore the microbial cell metabolic potential and to identify target genes for metabolic engineering. Depending on the desired purpose, the algorithm optimization methods can be divided into [42]: (1) linear programming; (2) quadratic programming; (3) mixed integer linear programming; and (4) evolutionary programming.

The most widespread method based on linear programming is Flux Balance Analysis (FBA) [43,44], which was specifically designed to predict quantitative growth phenotype based on a stoichiometric model and the pseudo-steady state assumption. FBA has been applied to (1) estimate the optimal state for growth under different cultivation conditions [45]; (2) maximize ATP or NADH production [46]; and (3) maximize the production of target metabolites [47,48]. In other words, FBA can predict global quantitative relationships between the input rates of nutrients [48], the output rates of byproducts and growth rates. Other algorithms based on linear programming are Flux Variability Analysis (FVA), which calculates the rank of possible variation of each reaction rate, Flux Coupling Analysis (FCA), which elucidates the correlation between different reaction rates, Thermodynamics based metabolic flux analysis, FBA with molecular crowding, etc. The Minimization of Metabolic Adjustment (MOMA) use quadratic programming (QP), and it has been developed to find a unique flux distribution that is the closest to a given flux distribution observed in a wild-type strain [49,50]. MOMA is based on the assumption that metabolism in a knockout mutant operates as closely as possible to metabolism in the wild-type strain. The algorithm has been proven to provide slightly better predictions than FBA.

In order to optimize the product yields of microbial strains, the most widespread strategy aims at coupling the production of the desired product to cellular growth. In order to achieve that, a number of optimization frameworks have been proposed. OptKnock, a bi-level optimization framework [51,52], was developed to identify optimal gene knockout strategies in order to obtain a resulting phenotype with a high production of the desired metabolite at the maximal growth rate. With OptKnock as starting point, four different computational approaches were proposed: OptReg

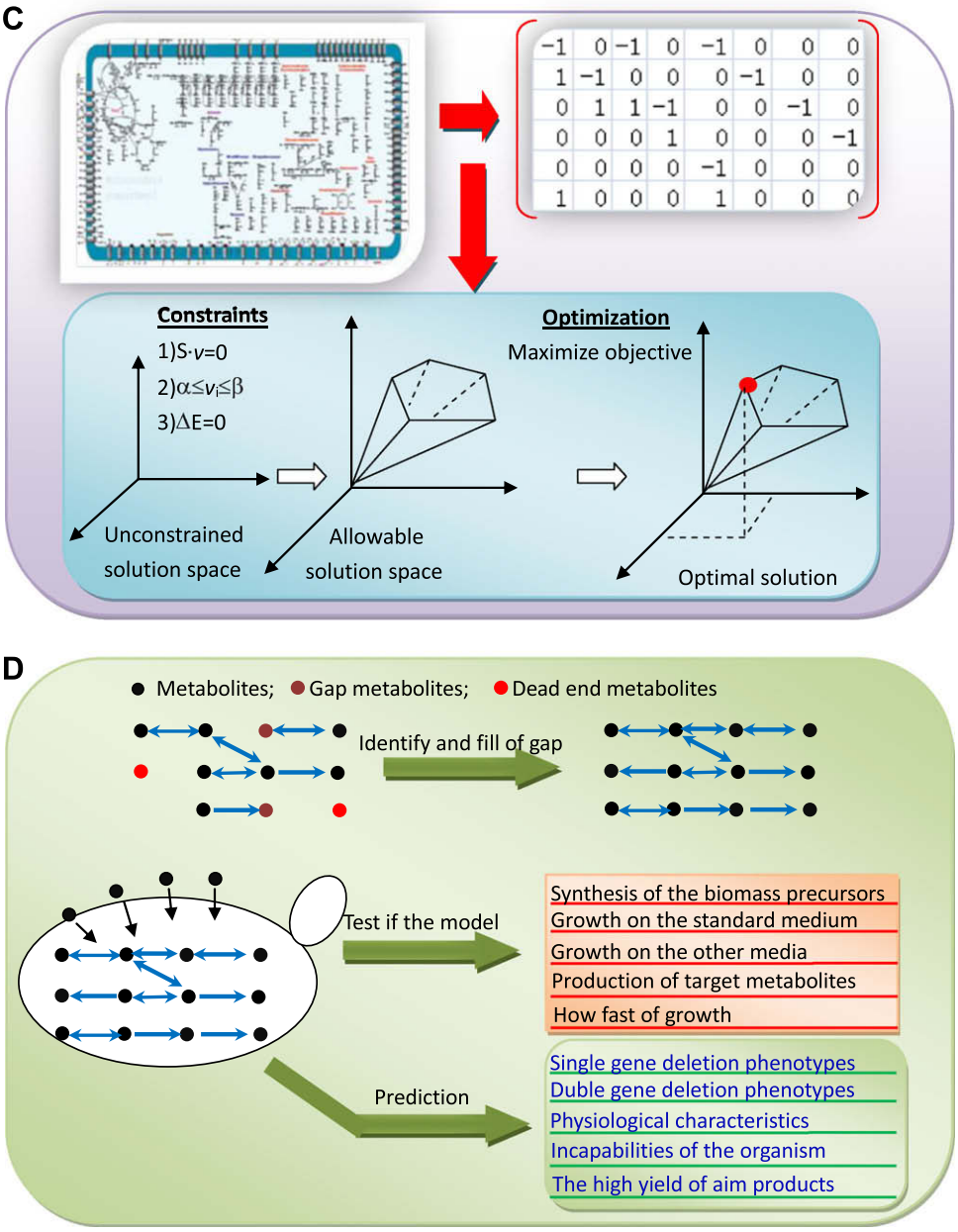

Fig. 3 (*continued*)

[53,54], OptStrain [55,56], OMNI (optimal metabolic network identification) [57,58], and regulatory on/off minimization (ROOM) [59]. OptReg initially uses constraints, such as glucose uptake rate, minimum ATP production and ^{13}C experimental flux data to characterize the upper and lower bounds of each reaction in the wild-type, secondly, by using a bi-level optimization it finds not only deletion targets but also over-expression and down-regulation targets in the form of changed upper and lower bounds. OptStrain identifies (among a collection of databases) a minimal set of non-native reactions that, when added to the wild strain, maximize the yield of the desired product. OMNI was designed to find the bottleneck reactions that cause the discrepancies between the experimental fluxes measured using ^{13}C labeling and in silico predictions. Similar to MOMA, ROOM also determines putative flux distributions after gene deletions by minimizing the number of significant flux changes [60,61]. OptGene [56] is an evolutionary algorithm developed to find deletion targets in microorganisms at lower computational costs. OptGene defines a priori the number of gene knockouts to be found and uses simulated annealing (SA) [62,63] and Set-based Evolutionary Algorithms (SEA) [62]. Each of the mentioned algorithms has its own particularities and are often best suited for specific cases as discussed by Park et al. [42].

4. Understanding of microbial physiology using genome-scale metabolic models

One of the primary uses of genome-scale metabolic models is the investigation of the effects of gene deletions [64,65]. Each reaction is associated to its enzyme-encoding genes, so gene deletions are translated into inactivated reactions with fluxes constrained to zero [66]. Following this rule, FBA [43,67], MOMA [49] and ROOM [59] methods have been developed and used to predict the outcomes of gene deletions series of bacteria. A main application of

GSMM is the identification of essential genes [68,69]. The essential genes are those which are required for a microbial cell survival [70]. Essential genes in a microorganism constitute its minimal gene set, which is the smallest possible group of genes that would be sufficient to sustain a functioning cell under the most favorable condition [69,71,72]. By using the most recent *E. coli* metabolic model, which contains 932 unique metabolic reactions, Andrew et al. tested 3888 single-deletion mutants on glycerol-supplemented minimal medium and found that 119 mutants were unable to grow on glycerol [73]. Furthermore, the determination of essential genes involved in the production of biomass precursors can be used to define relations between essentiality and particular metabolites [74]. Essentiality is in many cases environmental dependent [75,76]. Genome-scale metabolic models allow testing for essentiality in different growth conditions [77,78] in order to investigate the causes of gene dispensability, or to study bacterial evolution. Genome-scale metabolic models can also be used as frameworks for the interpretation of metabolic concentrations [79]. Differences in metabolite concentrations under known environmental conditions can be mapped onto genome scale models and combined with transcriptomic data to investigate the effects of metabolic regulation in the cell. Integrated with data on reaction directionality and standard chemical potentials, genome scale models have allowed the prediction of the concentration ranges of unmeasured metabolites, as well as the change in Gibbs free energy associated with each reaction and the identification of possible thermodynamic bottlenecks [80]. In addition to this, metabolic relationships between metabolites, either by simply examining the co-occurrence of metabolites in reactions, or by determining conservation relations between metabolites can also be elucidated [79,81,82]. At the level of metabolic fluxes, the relative occurrence of reactions within the total set of possible flux distributions [83] and across several environmental conditions has been evaluated by sampling methods [66,67], and the results showed that a few active reactions act as high-flux metabolic backbone while most of the others carry low fluxes. Similarly, the correlation and dependencies of pairs of metabolic reactions were determined [84,85]. The regulation of metabolism is a very complex process that involves gene–gene, gene–protein, protein–protein and protein–metabolite interactions [21]. There are many examples of genome scale models being used in conjunction with different high-throughput data in order to understand the regulation of metabolism. Recently, the role of transcriptional regulation of *Saccharomyces cerevisiae* in controlling metabolism was highlighted by comparing gene expression data and metabolic fluxes at a steady state in chemostat on various carbon sources. A remarkable regulatory-metabolic network of *S. cerevisiae* GSMM was reconstructed, and this include 55 regulatory transcription factors regulating 348 metabolic genes [86]. In *E. coli*, based on metabolic flux analysis, Perrenoud and Sauer investigated the impact of global transcription regulation by global regulatory proteins on glucose catabolism [87]. The regulatory role of the catabolite repressor/activator protein *cra* on the central metabolic pathways of *E. coli* was carefully investigated by analyzing metabolic fluxes and gene expression [88,89]. After the deletion of the *csrA* gene in *E. coli*, the formation of phosphoenolpyruvate, an essential precursor for synthesis amino acid, was inhibited by the repression of *pckA* and *pps* [90], therefore, it is not surprising to find that the phenylalanine titer increased with the over-expression *pps* gene in *E. coli* [91]. Similarly, disrupting three repressors of galactose uptake in *S. cerevisiae* can increase galactose consumption [92]. A comparison of in silico metabolic fluxes versus microarray gene expression data in *E. coli* and *S. cerevisiae* revealed that metabolic genes whose fluxes are directionally coupled generally share transcriptional regulators [93]. Comparing in silico predicted fluxes with expression states of metabolic genes, the most sensitive parts of metabolism to different stimulus such as carbon source shifts or nutrient limitations can be identified [94].

5. Concluding remarks and future directions

In order to enhance the performance of cell factories, a complete and comprehensive understanding of microbial physiology is necessary to guide metabolic engineering strategies. The increasing availability of high-throughput biological data allows us to understand cellular physiology at the genomic level. Genome-scale metabolic models include information on biochemical reactions that makes it possible to quantify the cell's metabolic capability and thus allow researchers to formulate predictions and make hypotheses that guide strategies for stain improvement. However, due to the lack of regulatory and kinetic information, the understanding of physiology attainable from genome-scale metabolic models is still limited. There is therefore a need for developing novel genome scale regulatory networks that integrate regulatory mechanisms with metabolism. Such integrative models are expected to appear in the near future and this will open for further advancement of computerized prediction of cellular phenotypes, and hence speed up the development of efficient cell factories that can be used for production of fuels, chemicals, feeds, foods and pharmaceuticals.

References

[1] Otero, J.M. and Nielsen, J. (2009) Industrial systems biology. Biotechnol. Bioeng. 105, 439–460.
[2] Park, J.H. and Lee, S.Y. (2008) Towards systems metabolic engineering of microorganisms for amino acid production. Curr. Opin. Biotech. 19, 454–460.
[3] Lin, H., Bennett, G.N. and San, K.Y. (2005) Fed-batch culture of a metabolically engineered *Escherichia coli* strain designed for high-level succinate production and yield under aerobic conditions. Biotechnol. Bioeng. 90, 775–779.
[4] Wendisch, V.F., Bott, M. and Eikmanns, B.J. (2006) Metabolic engineering of *Escherichia coli* and *Corynebacterium glutamicum* for biotechnological production of organic acids and amino acids. Curr. Opin. Microbiol. 9, 268–274.
[5] Atsumi, S. and Liao, J.C. (2008) Metabolic engineering for advanced biofuels production from *Escherichia coli*. Curr. Opin. Biotech. 19, 414–419.
[6] Chartrain, M., Salmon, P.M., Robinson, D.K. and Buckland, B.C. (2000) Metabolic engineering and directed evolution for the production of pharmaceuticals. Curr. Opin. Biotech. 11, 209–214.
[7] Liu, L.M., Li, Y., Li, H.Z. and Chen, J. (2006) Significant increase of glycolytic flux in *Torulopsis glabrata* by inhibition of oxidative phosphorylation. FEMS Yeast Res. 6, 1117–1129.
[8] Wright, J.A. et al. (2009) Multiple redundant stress resistance mechanisms are induced in *Salmonella enterica* serovar Typhimurium in response to alteration of the intracellular environment via TLR4 signalling. Microbiol-SGM 155, 2919–2929.
[9] Ho, S.N. (2003) The role of NFAT5/TonEBP in establishing an optimal intracellular environment. Arch. Biochem. Biophys. 413, 151–157.
[10] Jordan, S., Hutchings, M.I. and Mascher, T. (2008) Cell envelope stress response in Gram-positive bacteria. FEMS Microbiol. Rev. 32, 107–146.
[11] Ruiz, N. and Silhavy, T.J. (2005) Sensing external stress: watchdogs of the *Escherichia coli* cell envelope. Curr. Opin. Microbiol. 8, 122–126.
[12] Patil, K.R., Akesson, M. and Nielsen, J. (2004) Use of genome-scale microbial models for metabolic engineering. Curr. Opin. Biotech. 15, 64–69.
[13] Feist, A.M. and Palsson, B.O. (2008) The growing scope of applications of genome-scale metabolic reconstructions using *Escherichia coli*. Nat. Biotechnol. 26, 659–667.
[14] Ellis, R.J. (2001) Macromolecular crowding: an important but neglected aspect of the intracellular environment (vol. 11, p. 114, 2001). Curr. Opin. Struc. Biol. 11, 500.
[15] Begley, T.P., Chatterjee, A., Hanes, J.W., Hazra, A. and Lick, S.E.E. (2008) Cofactor biosynthesis – still yielding fascinating new biological chemistry. Curr. Opin. Chem. Biol. 12, 118–125.
[16] Brandt, U. (2006) Energy converting NADH : quinone oxidoreductase (Complex I). Ann. Rev. Biochem. 75, 69–92.
[17] Ponte-Sucre, A. (2007) Availability and applications of ATP-binding cassette (ABC) transporter blockers. Appl. Microbiol. Biotechnol. 76, 279–286.
[18] Davidson, A.L. (2002) Mechanism of coupling of transport to hydrolysis in bacterial ATP-binding cassette transporters. J. Bacteriol. 184, 1225–1233.
[19] Osley, M.A., Tsukuda, T. and Nickoloff, J.A. (2007) ATP-dependent chromatin remodeling factors and DNA damage repair. Mutat. Res. – Fun. Mol. M 618, 65–80.
[20] Xia, W.L. et al. (2009) Roles of NAD(+)/NADH and NADP(+)/NADPH in cell death. Curr. Pharm. Des. 15, 12–19.

[21] Vemuri, G.N. and Aristidou, A.A. (2005) Metabolic engineering in the -omics era: elucidating and modulating regulatory networks. Microbiol. Mol. Biol. Rev. 69, 197–216.

[22] Farmer, W.R. and Liao, J.C. (2000) Improving lycopene production in *Escherichia coli* by engineering metabolic control. Nat. Biotechnol. 18, 533–537.

[23] Turcotte, B., Liang, X.B., Robert, F. and Soontorngun, N. (2010) Transcriptional regulation of nonfermentable carbon utilization in budding yeast. FEMS Yeast Res. 10, 2–13.

[24] Kumar, P.R., Yu, Y., Sternglanz, R., Johnston, S.A. and Joshua-Tor, L. (2008) NADP regulates the yeast GAL induction system. Science 319, 1090–1092.

[25] Hynes, M.J., Murray, S.L., Duncan, A., Khew, G.S. and Davis, M.A. (2006) Regulatory genes controlling fatty acid catabolism and peroxisomal functions in the filamentous fungus *Aspergillus nidulans*. Eukaryot. Cell 5, 794–805.

[26] Hinnebusch, A.G. (2005) Translational regulation of GCN4 and the general amino acid control of yeast. Ann. Rev. Microbiol. 59, 407–450.

[27] Hube, B. (2009) Fungal adaptation to the host environment. Curr. Opin. Microbiol. 12, 347–349.

[28] Rudra, D., Zhao, Y. and Warner, J.R. (2005) Central role of Ifh1p–Fhl1p interaction in the synthesis of yeast ribosomal proteins. EMBO J. 24, 533–542.

[29] Rebholz-Schuhmann, D., Kirsch, H., Arregui, M., Gaudan, S., Riethoven, M. and Stoehr, P. (2007) EBIMed – text crunching to gather facts for proteins from Medline. Bioinformatics 23, E237–E244.

[30] Krallinger, M. and Valencia, A. (2005) Text-mining and information-retrieval services for molecular biology. Genome Biol. 6, 8.

[31] Spasic, I., Simeonidis, E., Messiha, H.L., Paton, N.W. and Kell, D.B. (2009) KiPar, a tool for systematic information retrieval regarding parameters for kinetic modelling of yeast metabolic pathways. Bioinformatics 25, 1404–1411.

[32] He, M., Wang, Y. and Li, W. (2009) PPI finder: a mining tool for human protein–protein interactions. Plos One 4, 6.

[33] Ananiadou, S., Kell, D.B. and Tsujii, J. (2006) Text mining and its potential applications in systems biology. Trends Biotech. 24, 571–579.

[34] Muller, H.M., Kenny, E.E. and Sternberg, P.W. (2004) Textpresso: an ontology-based information retrieval and extraction system for biological literature. Plos Biol. 2, 1984–1998.

[35] Goetz, T. and von der Lieth, C.W. (2005) PubFinder: a tool for improving retrieval rate of relevant PubMed abstracts. Nucleic Acids Res. 33, W774–W778.

[36] Becker, K.G., Hosack, D.A., Dennis, G., Lempicki, R.A., Bright, T.J., Cheadle, C. and Engel, J. (2003) PubMatrix: a tool for multiplex literature mining. BMC Bioinform. 4, 6.

[37] Maier, H., Dohr, S., Grote, K., O'Keeffe, S., Werner, T., de Angelis, M.H. and Schneider, R. (2005) LitMiner and WikiGene: identifying problem-related key players of gene regulation using publication abstracts. Nucleic Acids Res. 33, W779–W782.

[38] Dieterich, G., Karst, U., Wehland, J. and Jansch, L. (2005) MineBlast: a literature presentation service supporting protein annotation by data mining of BLAST results. Bioinformatics 21, 3450–3451.

[39] Thiele, I. and Palsson, B.O. (2010) A protocol for generating a high-quality genome-scale metabolic reconstruction. Nat. Protoc. 5, 93–121.

[40] Okuda, S., Yamada, T., Hamajima, M., Itoh, M., Katayama, T., Bork, P., Goto, S. and Kanehisa, M. (2008) KEGG Atlas mapping for global analysis of metabolic pathways. Nucleic Acids Res. 36, W423–W426.

[41] Chang, A., Scheer, M., Grote, A., Schomburg, I. and Schomburg, D. (2009) BRENDA, AMENDA and FRENDA the enzyme information system: new content and tools in 2009. Nucleic Acids Res. 37, D588–D592.

[42] Park, J.M., Kim, T.Y. and Lee, S.Y. (2009) Constraints-based genome-scale metabolic simulation for systems metabolic engineering. Biotechnol. Adv. 27, 979–988.

[43] Varma, A. and Palsson, B.O. (1994) Metabolic flux balancing-basic concepts, scientific and practical use. Bio-Technology 12, 994–998.

[44] Lee, J.M., Gianchandani, E.P. and Papin, J.A. (2006) Flux balance analysis in the era of metabolomics. Brief Bioinform. 7, 140–150.

[45] Raman, K. and Chandra, N. (2009) Flux balance analysis of biological systems: applications and challenges. Brief Bioinform. 10, 435–449.

[46] Ramakrishna, R., Edwards, J.S., McCulloch, A. and Palsson, B.O. (2001) Flux-balance analysis of mitochondrial energy metabolism: consequences of systemic stoichiometric constraints. Am. J. Physiol. – Reg. I 280, R695–R704.

[47] Lee, D.S., Burd, H., Liu, J.X., Almaas, E., Wiest, O., Barabasi, A.L., Oltvai, Z.N. and Kapatral, V. (2009) Comparative genome-scale metabolic reconstruction and flux balance analysis of multiple *Staphylococcus aureus* genomes identify novel antimicrobial drug targets. J. Bacteriol. 191, 4015–4024.

[48] Cakir, T., Efe, C., Dikicioglu, D., Hortacsu, A., Kirdar, B. and Oliver, S.G. (2007) Flux balance analysis of a genome-scale yeast model constrained by exometabolomic data allows metabolic system identification of genetically different strains. Biotechnol. Prog. 23, 320–326.

[49] Segre, D., Vitkup, D. and Church, G.M. (2002) Analysis of optimality in natural and perturbed metabolic networks. Proc. Natl. Acad. Sci. USA 99, 15112–15117.

[50] Rodrigues, J.F.M. and Wagner, A. (2009) Evolutionary plasticity and innovations in complex metabolic reaction networks. Plos Comput. Biol. 5, 11.

[51] Burgard, A.P., Pharkya, P. and Maranas, C.D. (2003) OptKnock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. Biotechnol. Bioeng. 84, 647–657.

[52] Pharkya, P., Burgard, A.P. and Maranas, C.D. (2003) Exploring the overproduction of amino acids using the bilevel optimization framework OptKnock. Biotechnol. Bioeng. 84, 887–899.

[53] Pharkya, P. and Maranas, C.D. (2006) An optimization framework for identifying reaction activation/inhibition or elimination candidates for overproduction in microbial systems. Metab. Eng. 8, 1–13.

[54] Fischer, E., Zamboni, N. and Sauer, U. (2004) High-throughput metabolic flux analysis based on gas chromatography–mass spectrometry derived C-13 constraints. Anal. Biochem. 325, 308–316.

[55] Pharkya, P., Burgard, A.P. and Maranas, C.D. (2004) OptStrain: a computational framework for redesign of microbial production systems. Genome Res. 14, 2367–2376.

[56] Patil, K.R., Rocha, I., Forster, J. and Nielsen, J. (2005) Evolutionary programming as a platform for in silico metabolic engineering. BMC Bioinform. 6, 12.

[57] Herrgard, M.J., Fong, S.S. and Palsson, B.O. (2006) Identification of genome-scale metabolic network models using experimentally measured flux profiles. Plos Comput. Biol. 2, 676–686.

[58] Fleming, R.M.T., Thiele, I. and Nasheuer, H.P. (2009) Quantitative assignment of reaction directionality in constraint-based models of metabolism: application to *Escherichia coli*. Biophys. Chem. 145, 47–56.

[59] Shlomi, T., Berkman, O. and Ruppin, E. (2005) Regulatory on/off minimization of metabolic flux changes after genetic perturbations. Proc. Natl. Acad. Sci. USA 102, 7695–7700.

[60] Nakahigashi, K. et al. (2009) Systematic phenome analysis of *Escherichia coli* multiple-knockout mutants reveals hidden reactions in central carbon metabolism. Mol. Syst. Biol. 5, 14.

[61] Zhao, Q.Y. and Kurata, H. (2009) Genetic modification of flux for flux prediction of mutants. Bioinformatics 25, 1702–1708.

[62] Rocha, M., Maia, P., Mendes, R., Pinto, J.P., Ferreira, E.C., Nielsen, J., Patil, K.R. and Rocha, I. (2008) Natural computation meta-heuristics for the in silico optimization of microbial strains. BMC Bioinform. 9, 16.

[63] Gonzalez, O.R., Kuper, C., Jung, K., Naval, P.C. and Mendoza, E. (2007) Parameter estimation using Simulated Annealing for S-system models of biochemical networks. Bioinformatics 23, 480–486.

[64] Imielinski, M., Belta, C., Halasz, A. and Rubin, H. (2005) Investigating metabolite essentiality through genome-scale analysis of *Escherichia coli* production capabilities. Bioinformatics 21, 2008–2016.

[65] Doerr, A. (2010) A global map of gene function. Nat. Meth. 7, 176.

[66] Durot, M., Bourguignon, P.Y. and Schachter, V. (2009) Genome-scale models of bacterial metabolism: reconstruction and applications. FEMS Microbiol. Rev. 33, 164–190.

[67] Price, N.D., Reed, J.L. and Palsson, B.O. (2004) Genome-scale models of microbial cells: evaluating the consequences of constraints. Nat. Rev. Microbiol. 2, 886–897.

[68] Hillyard, D.R. and Redd, M.J. (2007) Identification of essential genes in bacteriaAdvanced Bacterial Genetics: Use of Transposons and Phage for Genomic Engineering, pp. 21–34, Elsevier, San Diego.

[69] Gerdes, S., Edwards, R., Kubal, M., Fonstein, M., Stevens, R. and Osterman, A. (2006) Essential genes on metabolic maps. Curr. Opin. Biotech. 17, 448–456.

[70] Zhang, R. and Lin, Y. (2009) DEG 5.0, a database of essential genes in both prokaryotes and eukaryotes. Nucleic Acids Res. 37, D455–D458.

[71] Koonin, E.V. (2000) How many genes can make a cell: the minimal-gene-set concept. Annu. Rev. Genomics Hum. Genet. 1, 99–116.

[72] Manimaran, P., Hegde, S.R. and Mande, S.C. (2009) Prediction of conditional gene essentiality through graph theoretical analysis of genome-wide functional linkages. Mol. Biosyst. 5, 1936–1942.

[73] Joyce, A.R. et al. (2006) Experimental and computational assessment of conditionally essential genes in *Escherichia coli*. J. Bacteriol. 188, 8259–8271.

[74] Hara, K.Y., Shimodate, N., Ito, M., Baba, T. and Mori, H. (2009) Systematic genome-wide scanning for genes involved in ATP generation in *Escherichia coli*. Metab. Eng. 11, 1–7.

[75] Imielinski, M. and Belta, C. (2008) Exploiting the pathway structure of metabolism to reveal high-order epistasis. BMC Syst. Biol. 2, 15.

[76] Hong, C.C. and Song, M.Z. (2010) Optimal *in silico* target gene deletion through nonlinear programming for genetic engineering. Plos One 5, 9.

[77] Diamant, I., Eldar, Y.C., Rokhlenko, O., Ruppin, E. and Shlomi, T. (2009) A network-based method for predicting gene-nutrient interactions and its application to yeast amino-acid metabolism. Mol. Biosyst. 5, 1732–1739.

[78] Borodina, I., Krabben, P. and Nielsen, J. (2005) Genome-scale analysis of *Streptomyces coelicolor* A3(2) metabolism. Genome Res. 15, 820–829.

[79] Imielinski, M., Belta, C., Rubin, H. and Halasz, A. (2006) Systematic analysis of conservation relations in *Escherichia coli* genome-scale metabolic network reveals novel growth media. Biophys. J. 90, 2659–2672.

[80] Kummel, A., Panke, S. and Heinemann, M. (2006) Putative regulatory sites unraveled by network-embedded thermodynamic analysis of metabolome data. Mol. Syst. Biol. 2, 10.

[81] Becker, S.A., Price, N.D. and Palsson, B.O. (2006) Metabolite coupling in genome-scale metabolic networks. BMC Bioinform. 7, 14.

[82] Nikolaev, E.V., Burgard, A.P. and Maranas, C.D. (2005) Elucidation and structural analysis of conserved pools for genome-scale metabolic reconstructions. Biophys. J. 88, 37–49.

[83] Almaas, E., Kovacs, B., Vicsek, T., Oltvai, Z.N. and Barabasi, A.L. (2004) Global organization of metabolic fluxes in the bacterium *Escherichia coli*. Nature 427, 839–843.

[84] Reed, J.L. and Palsson, B.O. (2004) Genome-scale in silico models of *E. coli* have multiple equivalent phenotypic states: assessment of correlated reaction subsets that comprise network states. Genome Res. 14, 1797–1805.
[85] Becker, S.A., Feist, A.M., Mo, M.L., Hannum, G., Palsson, B.O. and Herrgard, M.J. (2007) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. Nat. Protoc. 2, 727–738.
[86] Herrgard, M.J., Lee, B.S., Portnoy, V. and Palsson, B.O. (2006) Integrated analysis of regulatory and metabolic networks reveals novel regulatory mechanisms in *Saccharomyces cerevisiae*. Genome Res. 16, 627–635.
[87] Perrenoud, A. and Sauer, U. (2005) Impact of global transcriptional regulation by ArcA, ArcB, Cra, Crp, Cya, Fnr, and Mlc on glucose catabolism in *Escherichia coli*. J. Bacteriol. 187, 3171–3179.
[88] Shimizu, K. (2009) Toward systematic metabolic engineering based on the analysis of metabolic regulation by the integration of different levels of information. Biochem. Eng. J. 46, 235–251.
[89] Saier, M.H. and Ramseier, T.M. (1996) The catabolite repressor/activator (Cra) protein of enteric bacteria. J. Bacteriol. 178, 3411–3417.
[90] Tatarko, M. and Romeo, T. (2001) Disruption of a global regulatory gene to enhance central carbon flux into phenylalanine biosynthesis in *Escherichia coli*. Curr. Microbiol. 43, 26–32.
[91] Wahl, A., El Massaoudi, M., Schipper, D., Wiechert, W. and Takors, R. (2004) Serial C-13-based flux analysis of an L-phenylalanine-producing *E. coli* strain using the sensor reactor. Biotechnol. Prog. 20, 706–714.
[92] Ostergaard, S., Olsson, L., Johnston, M. and Nielsen, J. (2000) Increasing galactose consumption by *Saccharomyces cerevisiae* through metabolic engineering of the GAL gene regulatory network. Nat. Biotechnol. 18, 1283–1286.
[93] Notebaart, R.A., Teusink, B., Siezen, R.J. and Papp, B. (2008) Co-regulation of metabolic genes is better explained by flux coupling than by network distance. Plos Comput. Biol. 4, 7.
[94] Usaite, R., Patil, K.R., Grotkjaer, T., Nielsen, J. and Regenberg, B. (2006) Global transcriptional and physiological responses of *Saccharomyces cerevisiae* to ammonium, L-alanine, or L-glutamine limitation. Appl. Environ. Microbiol. 72, 6194–6203.